Amendments to the Specification

Please amend the specification as follows.

Please amend paragraph [0002], at page 1, as follows:

[0002] There have been proposed a number of devices for once recording a sequence of events occurring at meetings, seminars, and interviews, communication over phones and videophones, images from televisions and monitor cameras, for later reproduction, by means of digital disks, digital still cameras, video tapes, or semiconductor memories, for example. The devices for such recording and reproduction have become popular as they are more ~~reliable,~~ reliable than hand writing, for recording sound and image information.

Please amend paragraph [0003], at pages 1-2, as follows:

[0003] With broadband communications that is recently widely available, information devices exemplarily including videophones, doorphones, and camera-equipped mobile terminals are now popularly used for person-to-person communication with sound and image information. For example, e-mails conventionally exchanged by text are now being replaced by videomails using sound and moving images. Also, with the widespread use of visualphones, messages left in answering machines so far recorded only by sound are now often accompanying video information. As such, simultaneous use of sound and moving images is now prevalent for the recent form of communication.

Please amend paragraph [0018], at pages 6-7, as follows:

[0018] Further, a predesignated face orientation determining step may determine whether or not the user is facing the front. A sound detection step may be also included to detect a sound included in the media. Moreover, the frame selecting step may select, by scanning the image sequence from the start point to the end point, and from the end point to the start point, the part of the image sequence ~~satisfying as being~~ between the time points determined in the determining step as the user facing the predesignated direction, and between time points at which a sound is each detected.

Please amend paragraph [0020], at page 7, as follows:

[0020] Further, the editing step may arrange ~~a~~ text included in the media onto an arrangement region or a speech bubble region which is so set as not to overlap at all the region extracted in the frame extracting step, or to overlap as little as possible if it overlaps.

Please amend paragraph [0023], at pages 8-10, as follows:

[0023] FIG. 1 is a block diagram showing the hardware structure of a media editing terminal capable of image communications realizing a media editing method of the present invention;

FIG. 2 is a block diagram showing the information flow and procedure of the processing at the time of media editing of the present invention;

FIG. 3 is a block diagram showing the functional structure of a media editing device according to a first embodiment;

FIG. 4 is a diagram for illustrating a clipping process applied to certain moving image data;

FIG. 5 is a diagram exemplarily showing meta data having index information of FIG. 4 described based on MPEG-7 standards;

FIG. 6 is a diagram showing an exemplary screen display of a terminal receiving a videomail which includes moving image data, and information (e.g., addresser, title);

FIG. 7 is a block diagram showing the functional structure of a media editing device according to a second embodiment;

FIG. 8 shows an exemplary trimming process and the resultant display screen;

FIG. 9 is a diagram showing exemplary meta-data Description for a partial region;

FIG. 10 shows an exemplary display screen showing only moving images with no space left for a title and a main text;

FIG. 11 shows an exemplary display screen where a title is arranged in a region not overlapping an image region including the user;

FIG. 12 shows an exemplary display screen where a main text is arranged in a region barely overlapping an image region including the user;

FIG. 13 is a diagram showing exemplary Description of meta data about a layout process of writing ~~a~~ text into moving images;

FIG. 14 shows an exemplary display image of a videomail on the receiver end having a character added;

FIG. 15 is a block diagram showing the functional structure of a media editing device according to a fourth embodiment;

FIG. 16 is a diagram exemplarily showing face characteristic values specifically focusing on the hair;

FIG. 17 is a diagram showing an exemplary editing screen for selecting which character to use;

FIG. 18 is a diagram showing an exemplary screen on the receiver end receiving a character mail;

FIG. 19 is a diagram showing another exemplary screen on the receiver end receiving a character mail; and

FIG. 20 is a block diagram showing the structure of a distributed-type media editing device (system).

Please amend paragraph [0025], at pages 10-11, as follows:

[0025] FIG. 1 is a block diagram showing the hardware structure of a media editing terminal where image communications ~~is~~ are carried out in such a manner as to realize the media editing method of the present invention. In FIG. 1, the present media editing terminal includes an input part 1, an image capturing part 2, an image display part 3, a sound input part 4, and a sound output part 5, all of which receive/provide information from/to the user. Further, included are an image-capturing control part 6, a sound input/output control part 7, a display control part 8, a communications part 9, a recording part 10, a recording control part 11, a signal processing part 12, and a control part 13, all of which process the information received/provided by the user. These constituents are interconnected via a system bus, an external bus, and the like. Here, the above structure is identical or similar to that of a general-type computer.

Please amend paragraph [0030], at page 12, as follows:

[0030] The sound output part 5 is composed of a speaker, and the like, and outputs, to the user, his/her recorded voice, received sound, and warning sound and beep as operationally necessary, for example.

Please amend paragraph [0036], at pages 13-14, as follows:

[0036] Here, the present media editing terminal may be of an integrated-type including every constituent mentioned above in one housing, or of a distributed-type performing data exchange among the constituents over a network or signal lines. For example, a camera-equipped mobile phone terminal is of the integrated-type carrying every constituent in a single housing. On the other hand, a doorphone is regarded as of the distributed-type because, at least, the image capturing part 2, the sound input part 4, and the sound output part 5 are externally located in the vicinity of the door, and the ~~remains~~ remaining parts are placed in another housing located in the living room, for example. This is for establishing an interface with visitors. Alternatively, such a distributed-type device may have a character database (later described) located outside.

Please amend paragraph [0041], at page 15, as follows:

[0041] Here, the clipping process and the layout process are both performed in the signal processing part 1, the control part 13, the recording control part 11, and the recording part 10 of FIG. 1. Typically, these processes are realized by a program executable by computer devices. The program is provided from a computer-readable recording medium, ~~e.g.~~ e.g., a CD-ROM, a semiconductor memory card, to the recording part 10, for example, and then downloaded over the communications lines.

Please amend paragraph [0044], at pages 16-17, as follows:

[0044] Generally, once the user creates a message in the form of videomail by his/her mobile terminal, he/she may have an itch to immediately send out the message. With the convenient interface provided, the user's ~~such~~ needs are thus met with a videomail created with a simple operation (e.g., one button operation). What is better, the resultant videomail layout is comprehensible to its addressee, having the message

clipped at the beginning and end, the image trimmed to have the user centered, and wallpaper and speech bubbles arranged as appropriate, for example. Herein, not all of the above processes are necessarily applied in the following embodiments, and combining any process needed for each different application will do. In the below, the embodiments of the present invention are individually described in detail.

Please amend paragraph [0056], at page 23, as follows:

[0056] Further, since the present media editing device performs both front determination and sound detection, clipping can be done with reliability to a part recorded as a message. Specifically, even if the user is facing the ~~camera~~ camera, but deep in thought, clipping never ~~miss~~ misses a time point when he/she starts speaking. Here, the present media editing device can achieve almost the same effects without sound detection. This is because the user normally faces toward the camera to start message recording, and thus front determination sufficiently serves the purpose. Also, if the user utters in spite of his/her intention before starting message recording, sound detection may not be considered effective. Therefore, the sound detection part 19 may be omissible.

Please amend paragraph [0057], at page 23, as follows:

[0057] Next, the editing part 21 performs media (moving image data) clipping on the basis of the starting and ending frames determined by the frame selection part 20. Here, the resultant moving image data generated by the editing part 21 may include only the clipped portion and ~~remains are all~~ the remainder is deleted, or the resultant data may be meta data including the clipped portion as an index. If the resultant data is meta data, no moving image data has been deleted, and thus any portion not ~~clipped~~ clipped, but important can be saved for later use. Exemplified below is a case where the meta data format is MPEG-7.

Please amend paragraph [0063], at pages 26-27, as follows:

[0063] As such, when the resultant data is meta data including a clipping portion as an index with no moving image data deleted, editing can be done without restraint if the data needs to be corrected after automatic clipping. This is because, unlike the case

where the resultant data is moving image data including only the clipped portion, ~~there needs to re-edit~~ only the meta data needs to be re-edited.

Please amend paragraph [0064], at pages 27-28, as follows:

[0064] In the above, the starting and ending frames provided by the frame selection part 20 are utilized for automatic clipping. Here, the starting frame may be defined as being an image appearing first on a terminal screen on the receiver end. In this sense, the clipping technique of the present media editing device is considered even more effective ~~even better~~. To be more specific, assuming a case where the user first sees a still image (e.g., a preview image, thumbnail image) showing what moving images are coming or already in storage. Here, such a still image is now referred to as an initial display image. In the example of FIG. 4, the first frame image is the one at point A. However, the image at A shows the user not facing towards the camera, and it is not considered suitable for the initial display image such as a preview or a thumbnail image. Accordingly, by using the meta data as illustrated in FIG. 5, the starting frame is defined as the initial display image. As a result, the frame image at point B showing the user facing the front is suitably displayed as the initial display image. The present media editing device thus has no need to newly transmit a still image as the initial display image to the receiver end. If newly transmitting, the media editing device uses the region extraction part 17 and the front determination part 18 to scan the data from the start point to the end point. Point B is resultantly detected, and the frame image corresponding thereto is transmitted as the initial display image. In this manner, the image showing the user facing the front appropriately goes to the receiver end.

Please amend paragraph [0068], at pages 28-29, as follows:

[0068] Described first is an assumable case in the present embodiment. Generally, any media to be transmitted in the form of videomail includes, not only moving image data, but information about who has sent the moving images with what title, for example. FIG. 6 is a diagram showing an exemplary screen display of a terminal receiving such a videomail. As shown in FIG. 6, on a display image 100, displayed are a moving image section 104, a header section 101 exemplarily indicating who has sent the

videomail to whom with what title, a text section 102, and a decoration section 103 having decorations appropriately laid out.

Please amend paragraph [0074], at page 30, as follows:

[0074] In FIG. 7, the basic data storage part 23 corresponds to the recording part 10 of FIG. 1, and stored therein are ~~such a~~ text as shown in FIG. 6, and basic data exemplified by image data for decoration. The layout part 22 reads, as appropriate, the basic data from the basic data storage part 23 by the user's operation, and performs the layout process including the trimming process. The details are left for later description.

Please amend paragraph [0075], at pages 30-31, as follows:

[0075] FIG. 8 shows an exemplary trimming process and the resultant display screen. In FIG. 8, shown in the upper portion is the moving image section 104 received from the same addresser of FIG. 6. Due to the reasons described in the above, the section contains a high proportion of background region behind the user's image. Thus, only the user region is trimmed in the following manner for laying out.

Please amend paragraph [0081], at pages 32-33, as follows:

[0081] When the meta data is used as such, unlike newly generating moving image data by cutting out a partial region therefrom, the amount of the moving image data is not reduced. The user on the receiver end, however, can freely change the layout according to the size of the terminal screen or his/her preference. For example, the user can relocate the partial region on the image to suit his/her preference, or make settings to display any other partial region. In such cases also, settings as the partial region set by the layout part 22 initially appearing on the screen ~~is~~ are considered convenient. This is because the region indicating who has sent the message is displayed first.

Please amend paragraph [0082], at page 33, as follows:

[0082] In MPEG-7, not only the method for setting "StillRegionDS" on a frame basis as shown in FIG. 9, but "MovingRegionDS" being information about any moving region, or "AudioVisualRegionDS" being information about a region with sound may be

used. As a comprehensive basic definition thereof, there is "SegmentDS" indicating a part of the multimedia contents. With any DS based on this definition, Description equivalent to that of FIG. 9 can be done with less ~~amount~~ space.

Please amend paragraph [0085], at pages 33-34, as follows:

[0085] Described first is an assumable case in the present embodiment, specifically a case where the display image 100 of FIG. 6 is trimmed in such a manner that the moving image section 104 occupies a larger ~~space~~ space, as much as ~~possible~~ possible, for display on a small screen (of mobile phone, for example). Here, presumably, information to be displayed on such a small screen is, at least, a "title", ~~a~~ "text", and moving images. Actually, the small screen is fully occupied only by the moving images, and there is no space left for the title and text. FIG. 10 shows an exemplary display screen showing only the moving images.

Please amend paragraph [0086], at page 34, as follows:

[0086] Here, the present media editing device is similar in structure to that of the second embodiment. To display such text information, however, the region extraction part 17 and the layout part 22 in the present media editing device are changed in their operations. In detail, onto the image region including the user (the user's image region) that has been detected by the region extraction part 17, the layout part 22 arranges the text information (e.g., title, text) so as not to overlap at all, or to overlap as little as possible if it overlaps. This operation is described in detail below.

Please amend paragraph [0087], at pages 34-35, as follows:

[0087] First, the region extraction part 17 detects the user's image region in the moving image data, and calculates the position and size thereof. Then, the layout part 22 receives the thus calculated position and size of the region, and the basic data (e.g., title, text) stored in the basic data storage part 23. The layout part 22 sets a region for arranging the ~~basis~~ basic data in the range not overlapping the user's image region at all (or overlapping as little as possible). FIG. 11 shows an exemplary display screen where a text title is arranged in a space not overlapping the user's image region. As shown in

FIG. 11, the text title is arranged in a space above the user's head with no overlap. With such arrangement, the resultant layout can contain any needed text together with moving images occupying a sizable proportion.

Please amend paragraph [0089], at pages 35-36, as follows:

[0089] The shape of the speech bubble region shown in FIG. 12 has, as quite familiar in cartoons, a sharp protrusion in the vicinity of the user's mouth. The position of the protrusion is calculated by an image recognition process. Specifically, the region extraction part 17 extracts a mouth region from the user's image region, and calculates its position. The layout part 22 arranges the protrusion onto the thus calculated position (or proximal position considered appropriate), and then sets the speech bubble region in the range not overlapping the user's image region at all (or overlapping as little as possible) in consideration of the number of letters of the text.

Please amend paragraph [0090], at page 36, as follows:

[0090] The resultant layout image is preferably displayed on the screen as the initial image (aforementioned initial display image) on the receiver end. That is, when opening incoming ~~mails~~ mail, the addressee first sees the image of FIG. 11 or 12, and checks only the title or the main text therewith. If the main text does not fit in one page, a scrolling process may be applied, for example. As such, the receiver checks a main text, for example, only in the first display ~~image~~ image, but not while the moving images are reproduced. This is surely not restrictive, and the main text or the title may be superimposed and displayed ~~during~~ when the moving images are reproduced so that the receiver can read the text while hearing and seeing the message in the form of the moving images.

Please amend paragraph [0093], at page 37, as follows:

[0093] Next, the layout part 22 preferably generates meta data which is the deciding factor for what layout in ~~the~~ a similar manner in the first and second embodiments. This is done to perform the layout process, that is, the process for writing a text into moving images.

Please amend paragraph [0100], at pages 39-40, as follows:

[0100] Described next is the operation of the media editing device of the present embodiment. The region extraction part 17 and the front determination part 18 operate in ~~the~~ a similar manner to the first embodiment, and determine whether or not the user in the moving images is facing the front. The result is forwarded to the editing part 26, from which any image determined as being the front image is provided to the character selection part 24. Based on thus received image(s), the character selection part 24 selects one or more of potential characters from the character database 25, where various ~~many~~ characters are stored as a database. Then, a character ID each corresponding to the thus selected character(s) are inputted into the editing part 26.

Please amend paragraph [0106], at pages 42-43, as follows:

[0106] In order to select any potential character registered in the character database 25 with reference to the extracted face characteristic values, used may be the aforementioned characteristic representations, or correlation values calculated with respect to the registered face characteristic values. Here, if the correlation value exceeds a threshold value set for the potential character images considered suitable, the corresponding character image is extracted as a potential. The character selection part 24 then notifies the character ID corresponding to the thus extracted potential character to the editing part 26.

Please amend paragraph [0110], at page 44, as follows:

[0110] FIG. 19 is a diagram showing another exemplary screen on the receiver end receiving the transmission data. As shown in FIG. 19, displayed in the lower part of the screen is a character selected by the user (addresser). Here, ~~during~~ when the message in the form of moving images is reproduced, the character may not be displayed, and in the meantime, the moving images may take over its display position. Such a layout may be generated by the editing part 26, or set on the receiver end.

Please amend paragraph [0111], at page 44, as follows:

[0111] Here, the number of potential ~~character~~ characters to be selected may be one, and if this is the case, mail creation becomes easier without selecting any potential character.

Please amend paragraph [0114], at pages 45-46, as follows:

[0114] In FIG. 20, such a distributed-type media editing device includes a character mail editing terminal 501, a character selection part 724, and a character database 725, which are interconnected over a network 600. Here, the character mail editing terminal 501 has the functions, partially or entirely, of the media editing devices of the first to third embodiments, and the character selection part 724 is located separately therefrom. Since this distributed-type media editing device is similar in structure and operation to the integrated-type, the same effects are to be achieved. Further, in the distributed-type media editing device of FIG. 20, in addition to the character mail editing terminal 501, the character selection part 724 and the character database 725 may be used also by a character mail reception terminal 502, or the like, where incoming mails are received and edited. If so, when receiving a character ID in ~~an~~ a character mail, the character mail reception terminal 502 only needs to receive the corresponding character image from the character database 725. In such a structure, terminals do not have to carry data large in amount. Moreover, in the case that the character mail reception terminal 502 operates as the media editing device when returning mails, the character selection part 724 and the character database 725 can be shared.

Please amend paragraph [0115], at page 46, as follows:

[0115] As such, in the distributed-type media editing device, the character selection part 724 and the character database 725 can be shared by a plurality of users. Therefore, terminals have no need to include those constituents, and can use databases storing various ~~many~~ characters.

Please amend paragraph [0116], at page 46, as follows:

[0116] As is known from the above, in the present media editing device, the user can easily create a character mail with any preferred character added thereto by narrowing down various ~~many~~ characters based on front images extracted from moving images. Further, with such a character mail, person-to-person communication can be smooth and active.